Combining Location and Classification Error Sources for Estimating Multi-Temporal Database Accuracy

Yohay Carmel, Denis J. Dean, and Curtis H. Flather

Abstract

Detection and quantification of temporal change in spatial objects is the subject of a growing number of studies. Much of the change shown in such studies may be an artifact of location error and classification error. The basic units of these two measures are different (distance units for location error and pixel counts for classification error). The lack of a single index summarizing both error sources poses a constraint on assessing and interpreting the apparent change. We present an error model that addresses location and classification error jointly. Our approach quantifies location accuracy in terms of thematic accuracy, using a simulation of the location error process. We further develop an error model that combines the location and classification accuracy matrices into a single matrix, representing the overall thematic accuracy in a single layer. The resulting time-specific matrices serve to derive indices for estimating the overall uncertainty in a multi-temporal dataset. In order to validate the model, we performed simulations in which known amounts of location and classification error were introduced into raster maps. Our error model estimates were highly accurate under a wide range of parameters tested. We applied the error model to a study of vegetation dynamics in California woodlands in order to explore its value for realistic assessment of change, and its potential to provide a means for quantifying the relative contributions of these two error sources.

Introduction

The detection and quantification of temporal change in spatial objects has been the subject of many studies during the last decade. Change-detection studies can be found within diverse disciplines such as geomorphology (Townsend and Walsh, 1998; Brown, 1999), landscape ecology (Turner and Gardner, 1992; Radeloff *et al.*, 2000), urban planning (Coulter *et al.*, 1999), and atmospheric sciences (Rees and Williams, 1997). These studies typically use remotely sensed datasets showing the same geographic region at different times to measure the extent of temporal change.

One common method for change detection is image differencing (e.g., Morisette *et al.*, 1999; Stow, 1999). In this analysis, the unaltered digital values of an image are subtracted from the digital values of another image of the same area. Another common method is post-classification (e.g., Carlson and Sanchez-Azofeifa, 1999; Carmel and Kadmon, 1999), in which change is assessed by means of a comparison of classified images or raster maps of an area. Change detection is one of the most successful applications of remote sensing (Singh, 1989), but it is subject to certain errors (Foody and Boyd, 1999). To date, no standard accuracy assessment technique for change detection has been developed (Congalton and Green, 1999, p. 80).

The goal of this paper is to develop and test an approach for assessing the accuracy of multi-temporal datasets accounting for different components of uncertainty. Our approach focuses on classified images and digital raster maps, where pixels represent nominal attributes.

Two major sources of error are associated with these types of spatial data sets: location error and classification error. Location error (sometimes called "spatial error," "positional error," or "misregistration") in the context of change detection is the misalignment between the various dataset layers. This type of error is typically quantified using root-mean-square error (RMSE) units (Brown, 1999). Classification error (often referred to as "attribute error" or "thematic error") is the error that arises from incorrect assignment of map classes to specific objects. Classification error is measured by counting the cases where actual and observed classes differ and is summarized in an accuracy matrix (also called an "error matrix"). Various indices of agreement between the classified image and some set of reference points can be derived from the accuracy matrix. These include the proportion of correctly classified (PCC) pixels, the user and producer accuracy, and the Kappa statistic (see review in Congalton and Green (1999), pp. 47–53). Note that classification error may include a location error component within it (i.e., errors caused by misregistration of the image to the ground). However, this location component of classification error should not be confused with the inter-image location error because, as it is defined here, location error refers only to the misregistration of the various maps in a time series. Of particular importance for change detection are two other sources of error (Jensen, 2000), radiometric effects (arising from attributes of the specific sensor being used) and atmospheric attenuation (the process whereby some of the energy of electromagnetic radiation is absorbed and/or scattered when traversing the atmosphere). Atmospheric conditions and sometimes the radiometer or its condition may change, and reduce the accu-

Y. Carmel is with the Faculty of Agricultural Engineering, The Technion-Israel Institute of Technology, Haifa, 32000 Israel (yohay@tx.technion.ac.il).

D.J. Dean is with the Department of Forest Sciences, Colorado State University, Fort Collins, CO 80523.

C.H. Flather is with the USDA Forest Service, Rocky Mountain Research Station, 2150 Center Ave., Bldg. A, Fort Collins, CO 80526-1891.

Photogrammetric Engineering & Remote Sensing
Vol. 67, No. 7, July 2001, pp. 865–872.

0099-1112/01/6707–865$3.00/0

racy of the observed change. In the case of post classification, these factors affect classification accuracy, and error of these sources is included within classification error.

In the case of multi-temporal datasets, where multiple data layers are involved (one for each point in time), much of the change shown in a change-detection analysis may in fact be an artifact of misregistration (Townshend *et al.*, 1992; Dai and Khorram, 1998; Stow, 1999) and misclassification (Cherrill and McClean, 1995). However, the basic units of these two measures of error are different (distance units for location error and pixel counts for classification error), and it is not possible to combine them directly into one overall measure of map quality (Lanter and Veregin, 1992). The lack of a single index summarizing both error sources poses a serious constraint to correctly assessing and interpreting the apparent change. In the present study, we derive an error model that jointly addresses location and classification error.

This paper includes three sections. The first section describes our error model for combining location and classification errors. Simulations we constructed to study and verify our model are illustrated in the second section. The third section explores an application of this model in a study of long-term vegetation changes in California woodlands, using a multi-temporal database.

The Error Model

Overview

Error analysis of multi-temporal datasets can be viewed as one form of error propagation analysis. In the last decade there has been a growing interest in the issue of error propagation in spatial data analysis. Error propagation models have been derived from error analysis theory (Taylor, 1982; Heuvelink, 1998), probability theory (Newcomer and Szajgin, 1984; Veregin, 1989; Veregin, 1994; Veregin, 1995), spatial statistics (Henebry, 1995), and Monte Carlo methods (Stoms *et al.*, 1992; Mowrer, 1997). However, only a few of these studies have explored the combined effects of location and classification errors. Arbia *et al.* (1998) used the corruption model (Geman and Geman, 1984) to assess the combined effects of these error sources. They used simulation and analysis-of-variance techniques to quantify the relative contributions of location and classification error components in simulated maps with known map properties (Arbia *et al.*, 1998). They found that, for a wide range of map properties and parameters tested, location error had a large effect on propagated error. However, possible interaction effects between error sources were not examined (Arbia *et al.*, 1998).

Each pixel in a multi-temporal database can be characterized by its specific set of temporal transitions. Let C_t denote the attribute value of a pixel in time t, where $t = 1, 2, ..., n$ time steps, and C can assume any one of $1, 2, ..., k$ nominal class values. $p(C_t)$ denotes the probability that pixel classification C in time period t is correct, i.e., the probability that a pixel assigned to category C actually belongs to category C. For a given pixel in a multi-temporal database, the probability that its assigned set of transitions indeed occurred is given by

$$p(C_1C_2 \ldots C_n) = p(C_1) \cap p(C_2) \cap \ldots \cap p(C_n). \quad (1)$$

Assuming independence of error between different time steps, Equation 1 becomes

$$p(C_1C_2 \ldots C_n) = p(C_1) \cdot p(C_2) \cdot \ldots \cdot p(C_n). \quad (2)$$

If registration of all layers was perfect, then each component of this index, $p(C_t)$, would be the user accuracy (Congalton and Green, 1999) associated with the attribute value C_t, derived from the appropriate classification accuracy matrix. However, misregistration of the various layers introduces an additional uncertainty component that needs to be accounted for by the error model.

We suggest a probabilistic approach for error analysis of multi-temporal datasets. It consists of a five-step process:

(1) Construct a location accuracy matrix, expressing the net impact of misregistration on image accuracy;
(2) Construct a classification accuracy matrix using the standard method of accuracy assessment;
(3) Combine both matrices to yield a single, overall accuracy matrix;
(4) Calculate the probability $p(C_t)$ for each attribute value C, as the user accuracy derived from the combined accuracy matrix; and
(5) Calculate the probability $p(C_1, C_2, ..., C_n)$ using Equation 2.

Construction of the Location Accuracy Matrix

Several studies addressed the impacts of location error on thematic accuracy (Townshend *et al.*, 1992; Dai and Khorram, 1998; Stow, 1999). Townshend *et al.* (1992) investigated those impacts using a simulation approach. A MODIS-N image was artificially mis-registered to itself by a known spatial lag. The resulting shifted image was compared to the original image and the "change"—the difference between both images—was recorded. They found that, even when misregistration is as small as a quarter of a pixel, its artifacts are likely to be greater than the actual change.

Here, we suggest a similar approach. An observed raster map $\mathbf{Q}_{observed}$ is shifted by x pixels horizontally and y pixels vertically, where x and y represent the horizontal and vertical components of the RMSE. The resulting map, $\mathbf{Q}_{shift}$, is compared to the original $\mathbf{Q}_{observed}$ on a pixel-by-pixel basis. A location accuracy matrix is constructed, $\mathbf{A}^{LOC} = \mathbf{Q}_{shift}$ by $\mathbf{Q}_{observed}$, where each matrix cell i, j stores the number of class j pixels assigned to class i due to this artificial shift. We assume that the matrix $\mathbf{A}^{LOC}$ is an estimate of the unknown (i.e., true) location accuracy matrix $\mathbf{Q}_{observed}$ by $\mathbf{Q}_{actual}$.

Construction of the Classification Matrix

Classification accuracy matrix is constructed using the standard accuracy assessment methods (Stehman and Czaplewski, 1998).

Combining the Two Accuracy Matrices

Our approach for combining location and classification error suggests a conditional sequence over which these two errors occur. We define location error as preceding classification error; thus, the latter is conditioned on the former (assuming that images are first rectified and then classified). In order to combine the two matrices, we traced the fate of pixels when location error and then classification error are introduced into a map. Table 1 illustrates three hypothetical matrices for a single map with three map categories. $\mathbf{A}^{LOC}$, $\mathbf{A}^{CLASS}$, and $\mathbf{A}^{BOTH}$ are the accuracy matrices for misregistration only, classification error only, and both error sources combined, respectively. Matrix cells contain pixel counts, where i indicates the row (representing the observed image) and j indicates the column (representing the actual image). The count n in cell i,j of $\mathbf{A}^{LOC}$ represents the number of class j pixels that were assigned to class i under misregistration. Similarly, $n_{i,j}$ in $\mathbf{A}^{CLASS}$ is the count of class j pixels that were assigned to class i under misclassification. The total number of actual class j pixels is denoted by n_{+j}.

An interpretation of $n_{i,j}$ in $\mathbf{A}^{BOTH}$ is more complicated. Consider $n_{1,2}$ in $\mathbf{A}^{BOTH}$—it represents the number of class 2 pixels that were assigned to class 1 due to the combined effect of both errors. This may result from any one of three scenarios:

- A class 2 pixel is assigned to class 1 due to misregistration ($\mathbf{A}^{LOC}_{1,2}$), thereafter correctly classified to class 1 ($\mathbf{A}^{CLASS}_{1,1}$). The number of pixels that would be subject to this process is therefore $\mathbf{A}^{LOC}_{1,2} * \mathbf{A}^{CLASS}_{1,1} / \mathbf{A}^{CLASS}_{n+1}$.

TABLE 1. HYPOTHETICAL ACCURACY MATRICES OF (A) LOCATION ERROR, (B) CLASSIFICATION ERROR, AND (C) COMBINED ERROR

a. $\mathbf{A}^{LOC}$

		Actual (j)		
		1	2	3
Observed (i)	1	$n_{(1,1)}$	$n_{(1,2)}$	$n_{(1,3)}$
	2	$n_{(2,1)}$	$n_{(2,2)}$	$n_{(2,3)}$
	3	$n_{(3,1)}$	$n_{(3,2)}$	$n_{(3,3)}$
Column totals	n_{+j}	n_{+1}	n_{+2}	n_{+3}

b. $\mathbf{A}^{CLASS}$

		Actual (j)		
		1	2	3
Observed (i)	1	$n_{(1,1)}$	$n_{(1,2)}$	$n_{(1,3)}$
	2	$n_{(2,1)}$	$n_{(2,2)}$	$n_{(2,3)}$
	3	$n_{(3,1)}$	$n_{(3,2)}$	$n_{(3,3)}$
Column totals	n_{+j}	n_{+1}	n_{+2}	n_{+3}

c. $\mathbf{A}^{BOTH}$

		Actual (j)		
		1	2	3
Observed (i)	1	$A_{1,1}^{LOC}*(A_{1,1}^{CLASS}/n_{+1}) + A_{2,1}^{LOC}*(A_{1,2}^{CLASS}/n_{+2}) + A_{3,1}^{LOC}*(A_{1,3}^{CLASS}/n_{+3})$	$A_{1,2}^{LOC}*(A_{1,1}^{CLASS}/n_{+1}) + A_{2,2}^{LOC}*(A_{1,2}^{CLASS}/n_{+2}) + A_{3,2}^{LOC}*(A_{1,3}^{CLASS}/n_{+3})$	$A_{1,3}^{LOC}*(A_{1,1}^{CLASS}/n_{+1}) + A_{2,3}^{LOC}*(A_{1,2}^{CLASS}/n_{+2}) + A_{3,3}^{LOC}*(A_{1,3}^{CLASS}/n_{+3})$
	2	$A_{1,1}^{LOC}*(A_{2,1}^{CLASS}/n_{+1}) + A_{2,1}^{LOC}*(A_{2,2}^{CLASS}/n_{+2}) + A_{3,1}^{LOC}*(A_{2,3}^{CLASS}/n_{+3})$	$A_{1,2}^{LOC}*(A_{2,1}^{CLASS}/n_{+1}) + A_{2,2}^{LOC}*(A_{2,2}^{CLASS}/n_{+2}) + A_{3,2}^{LOC}*(A_{2,3}^{CLASS}/n_{+3})$	$A_{1,3}^{LOC}*(A_{2,1}^{CLASS}/n_{+1}) + A_{2,3}^{LOC}*(A_{2,2}^{CLASS}/n_{+2}) + A_{3,3}^{LOC}*(A_{2,3}^{CLASS}/n_{+3})$
	3	$A_{1,1}^{LOC}*(A_{3,1}^{CLASS}/n_{+1}) + A_{2,1}^{LOC}*(A_{3,2}^{CLASS}/n_{+2}) + A_{3,1}^{LOC}*(A_{3,3}^{CLASS}/n_{+3})$	$A_{1,2}^{LOC}*(A_{3,1}^{CLASS}/n_{+1}) + A_{2,2}^{LOC}*(A_{3,2}^{CLASS}/n_{+2}) + A_{3,2}^{LOC}*(A_{3,3}^{CLASS}/n_{+3})$	$A_{1,3}^{LOC}*(A_{3,1}^{CLASS}/n_{+1}) + A_{2,3}^{LOC}*(A_{3,2}^{CLASS}/n_{+2}) + A_{3,3}^{LOC}*(A_{3,3}^{CLASS}/n_{+3})$

- A class 2 pixel is assigned correctly to class 2 in spite of misregistration ($\mathbf{A}_{2,2}^{LOC}$), thereafter assigned to class 1 due to classification error ($\mathbf{A}_{1,2}^{CLASS}$). The respective term for this event is $\mathbf{A}_{2,2}^{LOC} * \mathbf{A}_{1,2}^{CLASS} / \mathbf{A}_{n+2}^{CLASS}$.
- A class 2 pixel is assigned to class 3 due to location error ($\mathbf{A}_{3,2}^{LOC}$), thereafter assigned to class 1 due to classification error ($\mathbf{A}_{1,3}^{CLASS}$). This event is denoted by $\mathbf{A}_{3,2}^{LOC} * \mathbf{A}_{1,3}^{CLASS} / \mathbf{A}_{n+3}^{CLASS}$.

Thus, the value for ($\mathbf{A}_{1,2}^{BOTH}$) is denoted by

$$\mathbf{A}_{1,2}^{BOTH} = \mathbf{A}_{1,2}^{LOC} * \mathbf{A}_{1,1}^{CLASS} / \mathbf{A}_{n+1}^{CLASS} + \mathbf{A}_{2,2}^{LOC} * \mathbf{A}_{1,2}^{CLASS} / \mathbf{A}_{n+2}^{CLASS} + \mathbf{A}_{3,2}^{LOC} * \mathbf{A}_{1,3}^{CLASS} / \mathbf{A}_{n+3}^{CLASS}$$

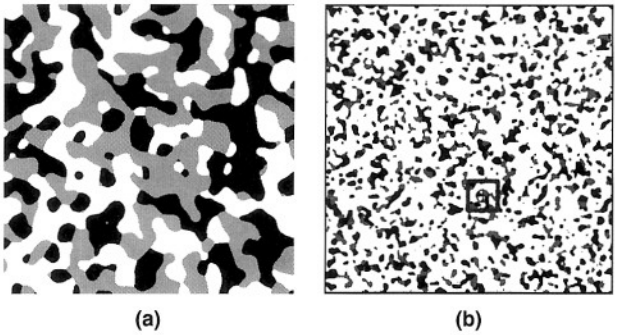
The expected counts for the remainder of the cells in the combined matrix $\mathbf{A}^{BOTH}$ are calculated similarly (Table 1c).

Simulating Uncertainty in a Multi-Temporal Database

A Single Map

A major theme in our error model is that location error and classification error interact in a predictable way. It follows that a straightforward validation of this model is possible. Consider the case where net effects of location and classification error in an image affected by both can be estimated separately; thus, the two respective error matrices may be constructed. Values in a combined error matrix, calculated using the equations in Table 1c, should approximate the values observed in the "actual" overall error matrix.

This error model is meaningful in the context of a multi-layer dataset. The goal of simulating error in a single map is to validate the model. We used the Arc-Info Grid module and Arc Macro Language (ESRI, 1996) for the simulations. A 512 by 512 thematic raster map was constructed, with one of k classes assigned randomly to each pixel, according to predetermined class proportions. A smoothing function was applied to the map in order to form a patchy pattern with a varying degree of clustering (Figures 1a and 1b). A location error map was constructed, shifting the original map by x pixels horizontally and y pixels vertically (Figure 2b). A classification error map was constructed by altering the pixel values for a given proportion of the original map (Figure 2c). Our algorithm allows the classification error to be patchy, similar to the original map or at any desired fragmentation level. The error rate for each class was controlled separately. A combined error map was constructed by first shifting the original image, and then introducing classi-



(a) (b)

Figure 1. Examples of simulated maps. (a) A clumped map (Moran I = 0.96) with equal class proportions. (b) A fragmented map (Moran I = 0.85) with unequal class proportions. The area within the central square is further enlarged in Figure 2.

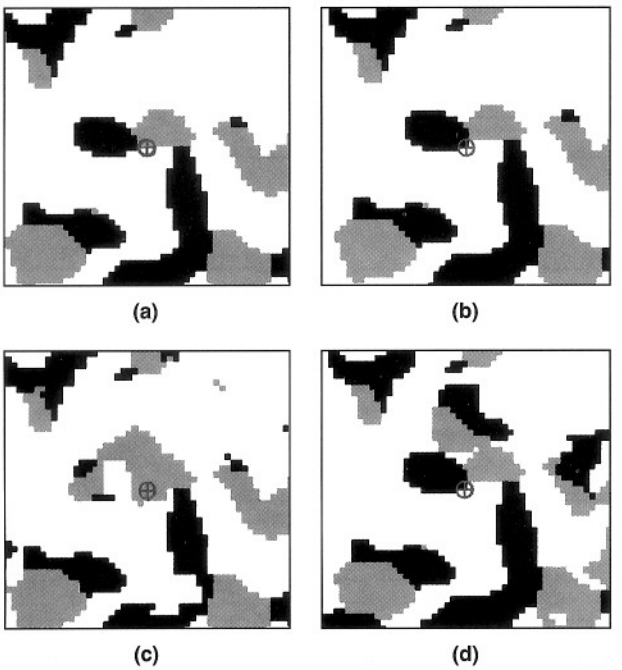


Figure 2. Small segments of simulated maps. (a) Original map. (b) The original map was shifted by 2 pixels up and 1 pixel left. (c) The original map was subject to 15 percent classification error. (d) The original map was shifted as in (b), then subjected to classification error similar to (c).

fication error (Figure 2d). Location, classification, and combined error matrices were derived from the respective maps. The observed combined error matrix $\mathbf{A}^{\text{OBS}}$ was compared to a predicted error matrix $\mathbf{A}^{\text{BOTH}}$, calculated using Table 1c equations. Differences between these two matrices were calculated as

$$D = \frac{\sum_{i=1}^{k}\sum_{j=1}^{k} |\mathbf{A}_{(i,j)}^{\text{OBS}} - \mathbf{A}_{(i,j)}^{\text{BOTH}}|}{N}$$

where N represents the total number of image pixels. We repeated this process in a set of simulations, testing a range of possible values for each of the following factors:

- Number of map categories (two to four).
- Proportions of each class in the original map (0.001 to 0.999).
- Degree of autocorrelation in the original map (Moran I values of 0.25 to 0.96).
- Degree of autocorrelation in classification error (Moran I values of 0.25 to 0.96).
- Location error rate (shift of 0.5 to 3 pixels).
- Classification error rate (PCC of 0.5 to 0.99).

Two Time Periods

The goal of this part of the simulation was to illustrate the potential use of our model in assessing the quality of a multitemporal dataset as a whole, and to determine the relative contribution of each error source. A thematic map with three classes for time period 1 (Map I) was constructed using the same method as described above. Pixel size was set to represent 10 m, the entire area representing 2600 ha. A second map for time period 2 (map II) was derived from the first map. Change was induced in the second map in a patchy manner. Location and classification error components were introduced to both maps in the same way as described above. Classification accuracy was set to be high (PCC = 0.93 and PCC = 0.95 in Map I and Map II, respectively). The overall misregistration between both maps was set to RMSE = 8 m. The combined accuracy matrices in the resulting maps were calculated using the same method as for the single map simulation. The observed change in each class was compared with the "actual" change. The probabilities associated with each transition type to be correct were calculated using Equation 2. The relative contribution of each error source to the accuracy associated with these transitions was measured.

Anticipating strong relations between map pattern and the impact of location error on thematic error, we ran this simulation for maps with two fragmentation patterns (see examples in Figure 1). The clumped and the fragmented patterns had average patch sizes of 2 ha and 0.2 ha, respectively.

Results

Results of the single map simulations confirmed our model predictions. The model was found to be robust under a wide range of the various parameters tested (Table 2). The cumulative differences between the matrices $\mathbf{A}^{\text{OBS}}$ and $\mathbf{A}^{\text{BOTH}}$ were less than 0.01 for all combinations of the tested parameters (Table 2). We therefore conclude that our model predicts correctly the combined effects of location and classification errors, and that differences between model predictions and observed results reflect stochasticity in the simulation process.

Results of simulating the process of change detection for two time periods demonstrate how the observed change may be very different from the real change, even when error for each source is very small. For example, for the fragmented and clumped images, 23 percent and 15 percent of class 1 in map I were found to change into other classes in map II, while the actual proportion for these transitions was 9 percent (Table 3). The probabilities $p(C_{\text{I}}^{1}C_{\text{II}}^{2})$ for an observed [class 1 − class 2] transition to be correct were 79 percent and 86 percent in the case of fragmented and clumped images, respectively (Table 3).

As expected, the degree of map fragmentation largely affected the degree of the overall error in the dataset. The overall PCC for the dataset, defined as the proportion of pixels correctly classified in both time periods, was 78 percent and 85 percent for the fragmented and clumped images, respectively.

Vegetation Change in Californian Woodlands: A Case Study

We applied our method to a case study in which the dynamics of oak woodlands have been studied over a period of 56 years. Four aerial photos of the Hastings Natural History Reservation and surroundings (Monterey County, California)—taken in 1939, 1956, 1971, and 1995—were scanned. Fourteen ground control points (GCPs) were identified in all four photos and measured in the field using a Magellan ProMARK X CM ™ GPS receiver. The RMSE of the ground control points was less than 0.6 m in all cases. Orthophotos were produced using the GCPs and a DEM of the area, using the ERDAS Imagine, OrthoBASE module. Location error for each of the four orthophotos was assessed using 40 points located across the scene that were identified in all photos. For a given point in a specific image, location error was defined as the Euclidean distance between the point location on that specific image and the average of locations of that point in all four images. The RMSE was calculated for each of the orthophotos.

The classification process followed the methods described by Carmel and Kadmon (1998). Basically, this is a hybrid of supervised/unsupervised classification, followed by a spatial filter. The classification scheme included three distinct vegetation classes: oak woodland, chaparral, and grassland (Griffin, 1977). Ground truthing was used to assess classification accuracy for the 1995 image, while a stereoscope-aided manual photointerpretation was used for the older images.

TABLE 2. RESULTS OF SOME SIMULATION RUNS FOR A SINGLE MAP. PCC IS THE PROPORTION OF PIXELS THAT WERE CLASSIFIED CORRECTLY. D IS THE CUMULATIVE ABSOLUTE DIFFERENCE BETWEEN PREDICTED AND OBSERVED MATRICES (SEE TEXT)

Parameter Tested (numbers refer to the list on p. 10) and characteristics	Class Proportions			PCC			D
	1	2	3	Location error only	Classification error only	Both error sources	Cumulative absolute difference
(2) One of the classes dominate	92%	4%	4%	0.96	0.89	0.89	7.67E-04
(2) One of the classes nearly absent	32%	68%	0.03%	0.86	0.88	0.77	2.04E-03
(5) Large location error	36%	33%	31%	0.61	0.89	0.56	3.91E-03
(6) Large classification error	30%	41%	30%	0.90	0.67	0.62	2.50E-03
(5,6) Large location and classification errors	34%	33%	34%	0.61	0.67	0.47	8.72E-03
(5,6) Moderate values for both error types	34%	33%	34%	0.90	0.92	0.83	1.50E-03
(3,5,6) Strong autocorrelation in the map[1], small values for both error types	34%	33%	33%	0.93	0.94	0.87	1.45E-03
(3,5,6) Strong autocorrelation in the map[1] moderate values for both error types	53%	38%	9%	0.87	0.83	0.73	2.02E-03
(3) Weak autocorrelation in the map[2]	34%	34%	31%	0.41	0.83	0.39	5.58E-03
(4) Strong autocorrelation in classification error pattern[3]	56%	22%	22%	0.87	0.86	0.76	2.82E-03
(4) Weak autocorrelation in classification error pattern[2]	34%	34%	31%	0.41	0.83	0.39	2.62E-03
(1) Two map classes	38%	62%		88%	84%	75%	1.71E-03
(1) Four map classes	25%	25%	25% 25%	92%	93%	85%	6.39E-03

[1]Moran's $I = 0.90$.
[2]Moran's $I = 0.29$.
[3]Moran's $I = 0.96$.

TABLE 3. RESULTS OF A SIMULATION OF CHANGE DETECTION FOR CLASS 1, SIMULATING TWO POINTS IN TIME. PIXEL SIZE IS 10 M. LOCATION ERROR: RMSE BETWEEN THE TWO MAPS IS 8 M. CLASSIFICATION ERROR: PCC IS 93% AND 95% FOR THE FORMER AND LATTER IMAGE, RESPECTIVELY

	Transition	'observed'	'actual'	$p(C_1C_j)$
Fragmented	1 to 1	0.77	0.91	0.798
Moran's I 0.81–0.85	1 to 2	0.10	0.03	0.792
	1 to 3	0.13	0.06	0.793
Clumped	1 to 1	0.85	0.91	0.853
Moran's I 0.92–0.96	1 to 2	0.03	0.001	0.861
	1 to 3	0.12	0.09	0.857

In order to construct location accuracy matrices, a shifted image was constructed for each image in the database, in which the shift on the x and y axes corresponded to the x and y RMSE values, respectively. The shifted images were compared to their precursors on a pixel-by-pixel basis, thus deriving the location accuracy matrices. The combined location-classification accuracy matrix was calculated for each time period.

Thematic accuracy for a multi-temporal database is usually described for two time steps, and its extension to n time steps deserves a brief discussion. Merging the various time-specific matrices using cross-tabulation was used by Congalton and Macleod (1994) for describing two time steps with two classes. This method has the advantage of presenting the entire accuracy information for the entire dataset. Yet, when more than two time steps are involved, the combined matrix becomes large. For example, the Hastings dataset accuracy assessment resulted in four matrices with three classes. Cross tabulation of these matrices would result in an 81 by 81 matrix. The interpretation of such matrices becomes difficult. Alternatively, the time-specific accuracy matrices may serve to construct other, more concise indices. We suggest three such indices, with an increasing degree of compaction. These are the transition-specific probability, the class-specific probability, and the spatio-temporal PCC.

The transition-specific probability, $p(C_{1939}C_{1956}C_{1971}C_{1995})$, is the probability that an observed transition is correct for all time steps. For example, the transition CGCC represents a pixel that was classified as chaparral in 1939, changed to grassland in 1956 (probably due to a fire event), changed to chaparral again in 1971, and remained classified chaparral in 1995. p(CGCC) is the probability that this exact transition indeed occurred. These probabilities were calculated as the products of the respective "user-accuracy" values for the relevant class in each time step.

The class-specific probability is the probability of any transition involving a specific map class i to be correct. It is calculated as the average of all the probabilities associated with transitions involving that specific class in at least one time step.

The spatio-temporal PCC is the probability that the assigned transition of any given pixel in the scene is correct. It can be calculated as the product of all PCCs derived from each time-step accuracy matrix.

Results

The classification accuracy for all images was relatively high; PCCs for the classification accuracy matrices were in the range of 0.90 to 0.94 (Table 4). Location accuracy was also relatively high (the RMSE was 3.53 m for 1939, 1.97 m for 1956, 2.42 m for 1971, and 2.51 m for 1995). However, location error had a large effect on the combined thematic accuracy (presumably due to the small pixel size, 0.6 m, relative to the RMSE, and due to the heterogeneous nature of the classified images). PCCs for the location accuracy matrices were 0.62 to 0.80 (Table 4). As a result, the combined accuracy PCCs were in the range of 0.58 to 0.74, being (as expected) lower than both their constituents (Table 4).

Table 5 presents three indices for describing the overall uncertainty in the dataset. For example, 43 percent of the 1939 grassland remained as such throughout the 65-year period; yet,

TABLE 4. ACCURACY MATRICES FOR HASTINGS VEGETATION MAPS

	Location			Classification			Combined		
1939	Forest	Chaparral	Grassland	Forest	Chaparral	Grassland	Forest	Chaparral	Grassland
Forest	4936095	111930	1393164	145	9	1	4284276	377535	1264901
Chap.	107710	2079572	236669	1	54	3	144655	1681490	344057
Grass.	1394644	232788	7105778	22	4	143	2009518	365265	7108991
PCC			0.80			0.90			0.74

	Location			Classification			Combined		
1956	Forest	Chaparral	Grassland	Forest	Chaparral	Grassland	Forest	Chaparral	Grassland
Forest	3785833	186933	2490121	126	2	4	3151134	214952	2186853
Chap.	197516	1160408	355944	4	38	3	316761	1014681	469368
Grass.	2440167	386870	4917592	25	4	144	2955621	504578	5123615
PCC			0.62			0.88			0.58

	Location			Classification			Combined		
1971	Forest	Chaparral	Grassland	Forest	Chaparral	Grassland	Forest	Chaparral	Grassland
Forest	5871814	153926	1585210	181	1	4	5426619	186040	1633273
Chap.	164847	1517133	268552	8	41	0	394425	1452788	320110
Grass.	1557050	283845	5026291	9	1	109	1772667	316076	4926670
PCC			0.76			0.94			0.72

	Location			Classification			Combined		
1995	Forest	Chaparral	Grassland	Forest	Chaparral	Grassland	Forest	Chaparral	Grassland
Forest	7260180	173030	1492576	169	3	2	6567349	269277	1475277
Chap.	177796	1328014	328503	5	29	1	345691	1048506	341749
Grass.	1457213	333164	4903500	14	5	107	1982149	516424	4925310
PCC			0.77			0.91			0.72

TABLE 5. INDICES FOR THE ACCURACY OF THE HASTINGS MULTI-TEMPORAL DATASET

(a) Transition Specific Accuracy, the Probability of a Specific Set of Transitions to Be Correct. Three Examples of Possible Transition Sets Are Presented.

Transition type	Nature of transition	Proportion of 1939 class undergoing this transition	Probability of being correct, accounting for: Location error	Classification error	Combined error
GGGG	Grassland does not change during the entire period	0.43	0.28	0.69	0.21
CGGG	Chaparral is burnt in the 1955 fire, thereafter is not recovering.	0.06	0.29	0.60	0.22
GTTT or GGTT or GGGT	Grassland turns into forest at some point during the 65 year period	0.37	0.31	0.69	0.25

(b) Class Specific Accuracy, the Probability for a Transition Involving a Particular Class to Be Correct, Averaged for all Relevant Transitions.

Transition type	Proportion of study area involved	Probability of being correct, accounting for: Location error	Classification error	Combined
All 69 transition combinations involving forest	0.66	0.26	0.59	0.20

(c) Spatio-Temporal PCC, the Probability for any Given Pixel, that Its Assigned Set of Transitions Is Correct, Averaged for all Pixels in the Scene.

Transition type	Proportion of study area involved	Probability of being correct, accounting for: Location error	Classification error	Combined
Averaged across the entire study area	1	0.29	0.67	0.22

for any single pixel, the probability that this transition sequence is correct is only 0.21 (Table 5a). Similarly, the combined probability of a transition sequence involving forest to be correct is 0.2 (Table 5b). The spatio-temporal PCC, which is the probability of any given pixel in this dataset being assigned to the correct class in all four time periods, is 0.22 (Table 5c). Note that, in all cases, the contribution of location error to the overall uncertainty is much larger than the contribution of clas-

sification error. This indicates that, for the Hastings dataset, reducing the impact of location error is a prerequisite before any meaningful analysis can take place. One way to increase spatial accuracy is by reducing resolution (Carmel *et al.*, 2001). Quantification of the gain in accuracy for a given decrease in resolution is the subject of an ongoing study in our lab.

Discussion

Location error between various layers in a GIS dataset may seriously affect the quality of any spatial analysis performed using these layers (Chrisman, 1989). This is particularly true in the case of change detection, where a time-series dataset is analyzed, as demonstrated here for the change in the Mediterranean woodlands of California. Existing methods of reporting thematic accuracy in multi-temporal datasets do not account for these location accuracy effects.

We suggest an approach that quantifies location accuracy in terms of thematic accuracy, using a simulation of the location error process. We further develop an error model that combines the location and classification accuracy matrices into a single matrix that represents the overall thematic accuracy for a single layer. In combining location and classification accuracy matrices, we make the assumption that, in the process of error propagation, location error precedes classification error. The interactions between both error sources are explicitly quantified in our model.

The resulting time-specific matrices can then serve to derive useful indices for estimating the overall uncertainty in a multi-temporal dataset. We suggest three such indices: the transition-specific probability, the class-specific probability, and the spatio-temporal PCC. Indices that account for the chance agreement between maps (e.g., kappa (Congalton, 1991)) may be derived from the time-specific combined matrices. Pontius (2000) suggests a specific derivation of kappa for maps with known location error. The extension of such indices for multi-temporal datasets is yet to be explored.

The simulations performed revealed that our error model estimates were found to be highly accurate under a wide range of parameters tested. These parameters include number and proportions of map classes, original map pattern, classification error pattern, and classification and location error rates.

Classification errors were found to be spatially correlated in previous studies (Congalton, 1988; Steele *et al.*, 1998). The effects of possible autocorrelation in classification error on the combined location-classification error were studied in the simulations. Our results show that autocorrelation in classification error did not affect the accuracy of model predictions for the combined error.

Yet, the robustness of our model to other aspects of error pattern was not tested, and further study may be needed. In particular, the effects of location error pattern, correlation between location and classification error patterns, and correlation between errors in different time steps should be investigated.

The location error induced in our simulations was a homogeneous shift. The error model assumes that the location accuracy matrix, $\mathbf{A}^{LOC}$, is an estimate of the unknown location accuracy matrix $\mathbf{Q}_{observed}$ by $\mathbf{Q}_{actual}$. However, location error in real maps is known to have complex spatial patterns (Brown, 1999). Further study is needed to assess if location error pattern affects the combined location-classification error.

In the process of combining both error sources, we let classification error be dependent on location error. Yet, the original probabilities for these two sources are assumed independent. It is difficult to assess the degree to which this assumption holds. In many applications, the data used for the classification and its accuracy assessment are collected independent of the data used in the orthorectification process. However, factors such an extremely steep terrain may affect both classification accuracy and location accuracy.

A third untested assumption of our model is independence of error between different time steps. Some correlation between error in different time steps should be expected for estimates of location accuracy (e.g., when the same DEM is used for rectification of all images, as in the case study presented here) and of classification accuracy (e.g., where classification errors may often be correlated with topography). The extent to which noncompliance with these assumptions biases model predictions is yet to be determined.

Acknowledgments

This study was funded by the USDA Forest Service, Inventory and Monitoring Institute, as part of the cooperative research program with the Keren Kayemet Leisrael (JNF). Thanks are due to Lucia Orlando of The Science Library of UC Santa Cruz and to Peter Ashley of HJW & Assoc. for their help with the air photos, and to Mark Stromberg of Hastings Nature Reserve.

References

Arbia, G., D. Griffith, and R. Haining, 1998. Error propagation modelling in raster GIS: Overlay operations, *International Journal of Geographical Information Systems*, 12:145–167.

Brown, D.G., 1999. Digital photogrammetric change analysis as applied to active coastal dunes in Michigan, *Photogrammetric Engineering & Remote Sensing*, 65:467–474.

Carlson, T.N., and G.A. Sanchez-Azofeifa, 1999. Satellite remote sensing of land use changes in and around San Jose, Costa Rica, *Remote Sensing of Environment*, 70:247–256.

Carmel, Y., and R. Kadmon, 1998. Computerized classification of Mediterranean vegetation using panchromatic aerial photographs, *Journal of Vegetation Science*, 9:445–454.

———, 1999. Grazing, topography, and long-term vegetation changes in a Mediterranean ecosystem, *Plant Ecology*, 145:239–250.

Carmel, Y., R. Kadmon, and R. Nirel, 2001. Spatio-temporal predictive models of Mediterranean vegetation dynamics, *Ecological Applications*, 11:268–280.

Cherrill, A., and C. McClean, 1995. An investigation of uncertainty in field habitat mapping and the implications for detecting land cover change, *Landscape Ecology*, 10:5–21.

Chrisman, N.R., 1989. Modeling error in overlaid categorical maps, *Accuracy of Spatial Databases*, (M.F. Goodchild and S. Gopal, editors), Taylor & Francis, London, pp. 21–34.

Congalton, R.G., 1988. Using spatial autocorrelation analysis to explore the errors in maps generated from remotely sensed data, *Photogrammetric Engineering & Remote Sensing*, 37:35–46.

———, 1991. A review of assessing the accuracy of classifications of remotely sensed data, *Remote Sensing of Environment*, 37:35–46.

Congalton, R.G., and K. Green, 1999. *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*, Lewis Publishers, New York, N.Y., 160 p.

Congalton, R.G., and R. Macleod, 1994. Change detection accuracy assessment on the NOAA Chesapeake Bay pilot study, *Proceedings of the International Symposium on Spatial Accuracy of Natural Resource Data Bases: Unlocking the Puzzle*, (R.G. Congalton, editor), May, Williamsburg, Virginia, American Society for Photogrammetry and Remote Sensing, Bethesda, Maryland, pp. 78–87.

Coulter, L., D. Stow, B. Kiracofe, C. Langevin, D.M. Chen, S. Daeschner, D. Service, and J. Kaiser, 1999. Deriving current land-use information for metropolitan transportation planning through integration of remotely sensed data and GIS, *Photogrammetric Engineering & Remote Sensing*, 65:1293–1300.

Dai, X.L., and S. Khorram, 1998. The effects of image misregistration on the accuracy of remotely sensed change detection, *IEEE Transactions on Geoscience and Remote Sensing*, 36:1566–1577.

ESRI, 1996. *Arc/Info 7.0.4 Users Guide*, Environmental Systems Research Institute, Redlands, California.

Foody, G.M., and D.S. Boyd, 1999. Detection of partial land cover change associated with the migration of inner-class transitional zones, *International Journal of Remote Sensing*, 20:2723–2740.

Geman, S., and D. Geman, 1984. Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images, *IEEE Transaction in Pattern Analysis and Machine Intelligence*, 6:721–742.

Griffin, J.R., 1977. Oak woodland, *Terrestrial Vegetation of California*, (M.G. Barbour and J. Major, editors), John Wiley & Sons, New York, N.Y., pp. 383–415.

Henebry, G.M., 1995. Spatial model error analysis using autocorrelation indices, *Ecological Modelling*, 82:75–91.

Heuvelink, G.B.M., 1998. *Error Propagation in Environmental Modelling with GIS*, Taylor and Francis, London, 144 p.

Jensen, J.R., 2000. *Remote Sensing of the Environment: An Earth Resource Perspective*, Prentice Hall, Upper Saddle River, New Jersey, 544 p.

Lanter, D.P., and H. Veregin, 1992. A research paradigm for propagating error in layer-based GIS, *Photogrammetric Engineering & Remote Sensing*, 58:825–833.

Morisette, J.T., S. Khorram, and T. Mace, 1999. Land-cover change detection enhanced with generalized linear models, *International Journal of Remote Sensing*, 20:2703–2721.

Mowrer, T.H., 1997. Propagating uncertainty through spatial estimation processes for old-growth subaplpine forests using sequential Gaussian simulation in GIS, *Ecological Modelling*, 98:73–86.

Newcomer, J.A., and J. Szajgin, 1984. Accumulation of thematic map errors in digital overlay analysis, *The American Cartographer*, 11:58–62.

Pontius, R.G., 2000. Quantification error versus location error in comparison of categorical maps, *Photogrammetric Engineering & Remote Sensing*, 66(8):1011–1016.

Radeloff, V.C., D.J. Mladenoff, and M.S. Boyce, 2000. Effects of interacting disturbances on landscape patterns: Budworm defoliation and salvage logging, *Ecological Applications*, 10:233–247.

Rees, W.G., and M. Williams, 1997. Monitoring changes in land cover induced by atmospheric pollution in the Kola Peninsula, Russia, using Landsat-MSS data, *International Journal of Remote Sensing*, 18:1703–1723.

Singh, A., 1989. Digital change detection techniques using remotely-sensed data, *International Journal of Remote Sensing*, 10:989–1003.

Steele, B.M., J.C. Winne, and R.L. Redmond, 1998. Estimation and mapping of misclassification probabilities for thematic land cover maps, *Remote Sensing of Environment*, 66:192–202.

Stehman, S.V., and R.L. Czaplewski, 1998. Design and analysis for thematic map accuracy assessment: Fundamental principles, *Remote Sensing of Environment*, 64:331–344.

Stoms, D.M., F.W. Davis, and C.B. Cogan, 1992. Sensitivity of wildlife habitat models to uncertainties in GIS data, *Photogrammetric Engineering & Remote Sensing*, 58:843–850.

Stow, D.A., 1999. Reducing the effects of misregistration on pixel-level change detection, *International Journal of Remote Sensing*, 20:2477–2483.

Taylor, J.R., 1982. *An Introduction to Error Analysis*, University Science Books, Mill Valley, California, 270 p.

Townsend, P.A., and S.J. Walsh, 1998. Modeling floodplain inundation using an integrated GIS with radar and optical remote sensing, *Geomorphology*, 21:295–312.

Townshend, J.R.G., C.O. Justice, C. Gurney, and J. McManus, 1992. The impact of misregistration on change detection, *IEEE Transaction on Geosience and Remote Sensing*, 30(5):1054–1060.

Turner, M.G., and R.H. Gardner (editors), 1992. *Quantitative Methods in Landscape Ecology*, Springer-Verlag, New York, N.Y., 520 p.

Veregin, H., 1989. Error modeling for map overlay, *Accuracy of Spatial Databases*, (M.F. Goodchild and S. Gopal, editors), Taylor and Francis, London, pp. 3–18.

———, 1994. Integration of simulation modeling and error propagation for the buffer operation in GIS, *Photogrammetric Engineering & Remote Sensing*, 60:427–435.

———, 1995. Developing and testing of an error propagation model for GIS overlay operations, *International Journal of Geographical Information Systems*, 9:595–619.

(Received 26 June 2000; accepted 27 October 2000; revised 05 December 2000)